B168

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C07K 14/35, C12N 15/31 — A2

(11) International Publication Number: WO 97/41252

(43) International Publication Date: 6 November 1997 (06.11.97)

(21) International Application Number: PCT/EP97/01973

(22) International Filing Date: 18 April 1997 (18.04.97)

(30) Priority Data:
196 17 184.9 29 April 1996 (29.04.96) DE

(71) Applicant *(for all designated States except US)*: GESELLSCHAFT FÜR BIOTECHNOLOGISCHE FORSCHUNG MBH (GBF) [DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SINGH, Mahavir [DE/DE]; (DE). HONISCH, Christiane [DE/DE]; (DE). ESPITIA, Clara [-/DE]; (DE). MORENO, Carlos [-/DE]; Mascheroder Weg 1, D-38124 Braunschweig (DE).

(74) Agents: BOETERS, Hans, D. et al.; Boeters & Bauer, Bereiteranger 15, D-81541 München (DE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DNA, RNA AND A PROTEIN USEFUL FOR DETECTION OF A MYCOBACTERIAL INFECTION

(57) Abstract

The invention concerns a DNA, RNA and a protein useful for identifying and combating mycobacterial infections.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	R	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Description

DNA, RNA AND A PROTEIN USEFUL FOR DETECTION OF A MYCOBACTERIAL INFECTION

Description

Technical field

The invention is in the field of clinical medicine, molecular biology and genetic engineering. More particularly, it relates to the molecular methods of tuberculosis diagnosis using newly identified DNA sequences which can be used as probes for DNA hybridization and or for DNA amplification leading to the identification of pathogenic mycobacteria causing disease in humans and animals.

Background

Tuberculosis, an infectious disease mainly caused by respiratory infection with *Mycobacterium tuberculosis*, represents an important subject of multidisciplinary investigation owing to the urgent need for rapid and reliable diagnostic tests and effective vaccines for disease control.

An estimated 8 million persons are developing tuberculosis each year and this number will be rising for the foreseeable future. Especially immuno-comprimised people, e.g. Human Immunodeficiency Virus-infected individuals (Selwyn et al., 1989; Barnes et al., 1991) and the population of countries with

insufficient public health systems (Grzybowski, 1991; Kochi, 1991) are the most endangered groups of this "global disease" (WHO, 1992). Emergence of multiple drug resistant strains is posing major threat to human health not only in developing countries, but also in developed countries. A rapid and specific diagnosis of tuberculosis is still a problem.

One approach to address this problem is to use the specific humoral or cellular response of the host to infer the presence of disease. Mycobacteria are rich in antigens that stimulate the production of antibodies and serology is simple and readily applicable as a rapid diagnostic test (Wilkins, 1994). Unfortunately the usefulness of serological tests are often limited by their lack of specificity and by their inability to destinguish between active disease, prior sensitization by contact with *M. tuberculosis* or cross-sensitization to other mycobacteria.

Another means of achieving the correct diagnosis are to develop increasingly sensitve methods to detect the causative bacilli or their products. Such techniques include amplification of a defined region of bacterial DNA via polymersase chain reaction (PCR) (Shankar et al., 1991), immunoassays for detecting antigen, gas liquid chromatography and mass spectrometry for detecting specific mycobacterial lipids. Of these, PCR is being evaluated most intensely and appears to hold greatest promise.

Attempts have been made to develop methods for the detection of chromosomal DNA of the *M. tuberculosis* complex in patient's sputum (Glennon, 1994). While the possibility of developing a DNA probe to distinguish between the *M. tuberculosis* complex and other mycobacterial strains has been reported, strain differentiation within the individual members of the complex is still a problem.

In this study we report the isolation of novel genomic clones containing as yet unreported genes and DNA, and the identification of novel *M. tuberculosis* chromosomal DNA regions specific for species of the *M. tuberculosis* complex. In addition, amplification of (i) a 377 bp fragment specific for the *M. tuberculosis* complex and (ii) of a 380-bp fragment showing sequence similarities with the genome of *Mycobacterium asiaticum*, *Mycobacterium gastri*, *Mycobacterium gordonae* and *Mycobacterium kansasii* are described. The utility of the 377-bp and the 380-bp fragment for the differentation of species and strains of mycobacteria is reported. In addition to other ORF identified in this study, a novel ca. 15kDa recombinant protein showing high homology to a family of transposase was overproduced in *Eschericha coli* as a thioredoxin fusion and purified. The ca. 15kDa and ca. 31kDa proteins described in this study are different from the 35kDa ORF belonging to an insertion element identified by Mariani et al. (1993).

Disclosure of invention

The present invention is based on novel DNA sequences cloned from the genome of *Mycobacterium tuberculosis,* which can be used for strain differentiation and for the diagnosis of tuberculosis.

Accordingly, the DNA sequences of the cloned fragments is an aspect of the invention.

The cloned DNA fragments are found to code for at least 7 proteins of about 9kDa, 15kDa, 17kDa,31kDa, 55kDa, 74kDa and 77kDa, the sequences of which are another aspect of the invention.

The use of the DNA sequence for detecting specific fragments by hybridization or by DNA amplification is another aspect of the invention.

The use of the cloned DNA or of the proteins coded by the cloned DNA for the purpose of serology, skin testing, vaccine development or drug design is another aspect of the invention.

The object underlaying the invention is solved by the following three main embodiments with their preferred embodiments.

According to a first embodiment the invention concerns a DNA
(a) having sequence (I) according to **figure** 9, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),
(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e).

Further the invention concerns a DNA according to (c) or (e), its single strands being hybridizable with those of the DNA according to (a), (b), (d) and (f), respectively, at a temperature of at least 25 ^{o}C and at a concentration of NaCl of 1 M.

Further the invention concerns a RNA being a transcript of a DNA according to the invention.

Further the invention concerns a protein being encoded by a DNA according to the invention.

Further the invention concerns a protein having the amino a. d sequence (II) according to **figure** 13.

The protein according to the invention can be an about 74 kDa protein.

Further the invention concerns a protein having the amino acid sequence (III) according to figure 14.

The protein according to the invention can be an about 77 kDa protein.

Further the invention concerns a protein having the amino acid sequence (IV) according to figure 15.

The protein according to the invention can be an about 9 kDa protein.

Further the invention concerns a protein having the amino acid sequence (V) according to figure 16.

The protein according to the invention can be an about 55 kDa protein.

The protein according to the invention can be a recombinant protein, especially a protein produced by means of a bacterial strain, a yeast strain, a fungal strain or a cell line of a higher eucaryote.

The protein according to the invention can be encoded by a DNA sequence according to the first embodiment of the invention and can be recovered by a method comprising the following steps:
(i) subjecting proteins encoded by said DNA sequence to a usual test for diagnosis of tuberculosis,
(ii) selecting a protein showing an inhibitory effect and
(iii) isolating and recovering said protein.

Further the invention concerns a DNA, RNA or protein according to the invention which can be used for
(i) diagnosis of tuberculosis in humans and animals and/or

(ii) diagnosis of other mycobacterial infections in humans or animals,
each especially by means of samples taken from humans or animals.

Further the invention concerns a use of a DNA according to the invention for the identification of mycobacteria in media samples.

The foregoing use can comprise the steps of
(i) isolating the mycobacterium,
(ii) preparing crude or purified genomic DNA,
(iii) hybridizing it to a DNA according to the first embodiment of the invention and
(iv) detecting the fragment pattern using conventional methods such as a radioactivity assay, chemiluminiscence or fluorescence.

Further the invention concerns a use, wherein as samples clinical samples are used.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for
(i) epidemeological purposes and/or
(ii) vaccination follow-up
for humans or animals suffering from mycobacterial infections, especially tuberculosis.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for the development of drugs useful for combating mycobacterial infections of humans or animals, especially tuberculosis, especially for testing and recovering of substances inhibiting mycobacterial infections in humans and animals, especially tuberculosis.

According to a second embodiment the invention concerns a DNA
(a) having sequence (VI) according to **figure 2**, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),
(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e).

Further the invention concerns a DNA according to (c) or (e), its single strands being hybridizable with those of the DNA according to (a), (b), (d) and (f), respectively, at a temperature of at least 25 ^{o}C and at a concentration of NaCl of 1 M.

Further the invention concerns a RNA being a transcript of a DNA according to the invention.

Further the invention concerns a protein being encoded by a DNA according to the invention.

Further the invention concerns a protein having the amino acid sequence (VII) according to **figure 5**.

The protein according to the invention can be an about 15 kDa protein.

Further the invention concerns a protein having the amino acid sequence (VIII) according to **figure 6**.

The protein according to the invention can be an about 31 kDa protein.

The protein according to the invention can be a recombinant protein, especially a protein produced by means of a bacterial strain, a yeast strain, a fungal strain or a cell line of a higher eucaryote.

The protein according to the invention can be a encoded by a DNA sequence according to the second embodiment of the invention and can be recovered by a method comprising the following steps:
(i) subjecting proteins encoded by said DNA sequence to a usual test for diagnosis of tuberculosis,
(ii) selecting a protein showing an inhibitory effect and
(iii) isolating and recovering said protein.

Further the invention concerns a DNA, RNA or protein according to the invention which can be used for
(i) diagnosis of tuberculosis in humans and animals and/or
(ii) diagnosis of other mycobacterial infections in humans or animals,
each especially by means of samples taken from humans or animals.

Further the invention concerns a use of a DNA according to the invention for the identification of mycobacteria in media samples.

The foregoing use can comprise the steps of
(i) isolating the mycobacterium,
(ii) preparing crude or purified genomic DNA,
(iii) hybridizing it to a DNA according to the second embodiment of the invention and

(iv) detecting the fragment pattern using conventional methods such as a radioactivity assay, chemiluminiscence or fluorescence.

Further the invention concerns a use, wherein as samples clinical samples are used.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for
(i) epidemeological purposes and/or
(ii) vaccination follow-up
for humans or animals suffering from mycobacterial infections, especially tuberculosis.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for the development of drugs useful for combating mycobacterial infections of humans or animals, especially tuberculosis, especially for testing and recovering of substances inhibiting mycobacterial infections in humans and animals, especially tuberculosis.

According to a third embodiment the invention concerns a DNA
(a) having sequence (IX) according to **figure** 3, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),
(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e).

Further the invention concerns a DNA according to (c) or (e), its single strands being hybridizable with those of the DNA

according to (a), (b), (d) and (f), respectively, at a temperature of at least 25 °C and at a concentration of NaCl of 1 M.

Further the invention concerns a RNA being a transcript of a DNA according to the invention.

Further the invention concerns a protein being encoded by a DNA according to the invention.

Further the invention concerns a protein having the amino acid sequence (X) according to figure 7.

The protein according to the invention can be an about 17 kDa protein.

The protein according to the invention can be a recombinant protein, especially a protein produced by means of a bacterial strain, a yeast strain, a fungicidal strain or a cell line of a higher eucaryote.

The protein according to the invention can be encoded by a DNA sequence according to the third embodiment of the invention and can be recovered by a method comprising the following steps:
(i) subjecting proteins encoded by said DNA sequence to a usual test for diagnosis of tuberculosis,
(ii) selecting a protein showing an inhibitory effect and
(iii) isolating and recovering said protein.

Further the invention concerns a DNA, RNA or protein according to the invention which can be used for
(i) diagnosis of tuberculosis in humans and animals and/or
(ii) diagnosis of other mycobacterial infections in humans or animals,

each especially by means of samples taken from humans or animals.

Further the invention concerns a use of a DNA according to the invention for the identification of mycobacteria in media samples.

The foregoing use can comprise the steps of
(i) isolating the mycobacterium,
(ii) preparing crude or purified genomic DNA,
(iii) hybridizing it to a DNA according to the third embodiment of the invention and
(iv) detecting the fragment pattern using conventional methods such as a radioactivity assay, chemiluminiscence or fluorescence.

Further the invention concerns a use, wherein as samples clinical samples are used.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for
(i) epidemeological purposes and/or
(ii) vaccination follow-up
for humans or animals suffering from mycobacterial infections, especially tuberculosis.

Further the invention concerns a use of a DNA according to the invention or of a protein according to the invention for the development of drugs useful for combating mycobacterial infections of humans or animals, especially tuberculosis, especially for testing and recovering of substances inhibiting mycobacterial infections in humans and animals, especially tuberculosis.

The invention is explained in detail by the following figures and experimental data.

Fig. 1 shows a restriction endonuclease map of the 7.2 kb *M. tuberculosis* chromosomal region;

Fig. 2. shows a 2253 bp *M. tuberculosis* chromosomal region including BamHI, EcoRI and KpnI restriction sites and oligonucleotides for screening the lambda gt 11 *M. tuberculosis* library (Primer 1 and Primer 2 underlined) and for amplification of the 377 bp region (377 bp region in bold, Primer 3 and Primer 4 underlined); amino acid sequences of the about 15 kDa and the about 31 kDa proteins are shown above the DNA sequences and are marked with arrows (small arrow about 15 kDa ORF 1, strong arrow about 31 kDa ORF 2);

Fig. 3. shows a DNA sequence of the 440 bp *M. tuberculosis* chromosomal region including the 380 bp region (in bold) used in PCR experiments and the amino acid sequence of the ORF 3 shown below the complementary DNA strand (< ORF 3);

Fig. 4 is an overview of the isolated lambda gt11-clone C9-2; 7.2 kb insert fragment, sequenced chromosomal regions and ORF 1, ORF 2 and ORF 3 marked with arrows;

Fig. 5 shows the amino acid sequence of the about 15 kDa protein (ORF 1);

Fig. 6 shows the amino acid sequence of the about 31 kDa protein (ORF 2);

Fig. 7 shows the amino acid sequence of the about 17 kDa protein;

Fig. 8 A shows SDS-PAGE of the insoluble pellet fraction (lane 1) and the purified about 15 kDa recombinant antigen (lane 2); lane A3 shows protein molecular weight standards (2.850 to 43.000 molecular weight range);

Fig 8 B shows SDS-PAGE of the purified about 15 kDa thioredoxin fusion protein (lane 1) and the two protein bands obtained after enterokinase cleavage (lane 1);

Fig. 9 shows a DNA sequence of *M. tuberculosis*;

Fig. 10 is a schematic drawing of the clone Mtub-Clara-Klon; the open reading frames of about 9 kDa (bp 3536 to bp 3829), 55 kDa (bp 2111 to bp 3829), 74 kDa (bp 1538 to bp 3829) and 77 kDa (bp 2698 to bp 2 on the complimentary strand) proteins are shown by arrows and the corresponding coding regions are numbered;

Fig. 11 A shows are southern hybridization with genomic DNA from different mycobacteria digested with PvuII (1: *M. tuberculosis* H37Rv; 2: *M. avium*; 3: *M. kanssasi*; 4: *M. necroti*; 5: *M. fortuitum*; 6: *M. phlei*; 7: *M. smegmatis*; 8: *M. vaccae*);

Fig. 11 B shows a finger-print obtained using the DNA (BamHI digest) of (1) *M. tuberculsosis* H37 RV, (2) *M. tuberculosis* H37 Ra, (3) *M. bovis* BCG, and (4) *M. tuberculosis* H37Rv digested with SalI;

Fig. 12 shows a finger-print with DNA from different *M. tuberculosis* clinical isolates (numbered 1 to 12) digested with PvuII restriction enzyme; the 4 kb Sal I fragment (Mtub-Klar-Klon) was used as probe;

Fig. 13 shows an amino acid sequence of the protein of about 74 kDa (molecular weight 74999, length 764)

Fig. 14 shows a glycine rich protein of about 77 kDa (molecular weight 77056, length 899);

Fig. 15 shows the amino acid sequence of the about 9 kDa proline rich protein (molecular weight 9356, length 98); and

Fig. 16 shows the proline rich protein of about 55 kDa (molecular weight 55982, length 573).

Modes for Carrying out the invention

We were interested in identifying and cloning novel DNA sequences from the genome of *Mycobacterium tuberculosis* for use in rapid and specific diagnosis of tuberculosis. Our strategy was to go for new repeated elements and insertion elements which are present only in *M.tuberculosis* or in the strains of *M. tuberculosis complex.*

Examples

The following examples further describe the isolation and sequencing of *M. tuberculosis*-DNA containing putative IS-element (Insertion Element) and repeat sequences, e.g., PGRS-elements (Polymorphic GC-Rich-Sequences) and the use of the as yet unreported DNA sequences for strain identification and diagnosis of tuberculosis.

Escherichia coli strains, phages and plasmids: The *Escherichia coli* K12 strain Y1090r$^-$ (Huynh et al., 1985) was used to propagate the λgt11 library and the *E. coli* K12 strain GI724 (Invitrogen, Leek, The Netherlands) was the host for the production of the ca. 15kDa protein fused to thioredoxin.

The recombinant DNA library of *M. tuberculosis* genomic DNA in the λgt11 expression vector was constructed by Young et al. (1985).

The plasmid vector pTrxFus (Invitrogen, Leek, The Netherlands) was used to make an in-frame fusion with thioredoxin as an amino-terminal fusion partner.

Mycobacterial strains and preparation of cell extracts: The mycobacterial strains used in this study are shown in Table 1 (Results and Discussion). All organisms

were grown on Loewenstein medium. For preparing cell extracts a loop of bacteria was suspended in 0.5 ml of 10 mM Tris/base, 1 mM EDTA (pH 7.4) followed by addition of 0.5 ml glass beads (150-212 microns, Sigma, Deisenhofen, Germany). The suspension was incubated at 80°C for 10 min followed by a 1 min treatment in a Mini-Bead Beater (Biospec Products).

DNA sequence analysis: Similarity comparisons were done using the BLAST program (Pearson and Lipman, 1988; NCBI computing facility).

All DNA manupulations were done according to standard procedures (see Maniatis et al. 1982).

DNA sequencing: DNA sequencing analysis was performed by the dideoxynucleotide-chain termination method using a PCR sequencing kit (ABI PRISM^{TM} Dye Terminator Cycle Sequencing Ready Reaction Kit, Perkin Elmer, Warrington, Great Britain) on a 373A DNA Sequencer (Applied Biosystems, Warrington, Great Britain). DNA sequences were determined for both strands by primer walking.

1. Clone containing putative IS-Element

1.1 **Isolation of the clone C9-2 containing a putative IS element:**

In our attempt to isolate new mycobacterial insertion elements, a λgt11 M. tuberculosis library was screened with oligodeoxyribonucleotide primers based on conserved regions of different insertion elements.The library was screened as described by Young and Davis (1985). Briefly, phage-infected cells of the strain E.coli Y1090r$^-$ were plated in top agar on Luria-Bertani plates (7.0×10^6 PFU per 85 mm plate) and incubated for 6-8 h at 42°C. Nylon membranes (Biodyne B Transfer Membrane, 0.45 µm, Pall, Portsmouth, England) were overlaid on plates. The filters were treated with 0.5 N NaOH, 1.5 M NaCl and the DNA was fixed via UV-crosslinking. Screening was performed using 3'-end labeled oligonucleotides of the sequence 5'-TGACGCGAGTGGGTGTGATTTCG-3' and 5'-GTGGTCGAGCCGTTGATGCCG-3' (Fig.2, PRIMER 1 and PRIMER 2) . Digoxigenin-labeling of the oligodeoxyribonucleotide primers was carried out using a DIG Oligonucleotide 3'-End Labeling Kit (Boehringer Mannheim, Germany). Hybridzation was done at 45°C in hybridization buffer (Boehringer Mannheim, Germany) overnight. Then the membranes were washed under stringent conditions for 5 min twice in 2 x SSC, 0.1% SDS and for 15 min twice at 37°C in 0.1 x SSC, 0.1% SDS. Chemi-luminescent detection was carried out with the help of a DIG Luminescent Detection Kit

(Boehringer Mannheim, Germany). Plaques were purified by three rounds of plating to obtain single plaques. Phage DNA was isolated using a Nucleobond AX L50 Kit (Machery-Nagel, Düren, Germany) and restriction mapping of the selected clone was performed by standard procedures (Maniatis et al., 1982).

Several positive clones were obtained. Detailed analysis of one of the clones (C9-2) is presented here. The recombinant phage was mapped with the restriction endonucleases BamHI, EcoRI and KpnI (Fig. 1). EcoRI digestion revealed a 7.2 kb DNA insert fragment.

1.2 **DNA sequencing of the cloned fragment:**

Two M. tuberculosis chromosomal regions of 2253-bp and 440-bp of this fragment were sequenced (Fig.2 and Fig.3). DNA sequencing of the 2253-bp region revealed the presence of a putative insertion element between bp 401 and bp 1378 containing inverted repeats flanked by duplications of 4 base pairs. The cloned fragment reported here is novel and is located at a different position than the 2.1 kb PstI/EcoRI fragment reported by Mariani et al. (1993), because the DNA sequence of the adjoining regions on the left and the right ends of the putative IS-element were completely different in our clone C9-2 as compared to that reported by Mariani et al. (1993).

Fig.4 gives an overview of the 7.2 kb insert fragment and the sequenced chromosomal regions.

1.3 **Novel Proteins coded by the cloned DNA:**

During the molecular characterization of the clone, novel ORFs were identified. The complete ORF of the ca. 15kDa protein is located on the 2253-bp fragment coded by a 408-bp fragment, corresponding to a coding capacity of 136 amino acids. The ca. 15kDa protein (Fig.5) is a novel product showing limited homology in the N-terminus of a 34kDa ORF reported by Mariani et al. (1993). We also identified an ORF of about 31kDa (Fig. 2 and Fig. 6) coded by the cloned DNA (bp 515 till bp 1378). This 31kDa ORF did not show any homology in the N-terminus to any known sequence in the database. The C-terminus of the ca. 31kDa protein showed homology to a 34kDa ORF (Mariani et al., 1993). We have not used the DNA sequence showing homology to the sequence reported by Mariani et al. (1993) as far as the claims of this patent application are concerned. An ORF (ORF 3, Fig. 3 and Fig.7) on the complementary strand to the 3'-end of the insert fragment of the recombinant λ-clone C9-2 was identified, which had not been reported earlier. This sequence showed homology to a family of transcription regulators in microorganism. In addition, some homology was observed with a

putative two-component system mtrA-mtrB isolated from M.tuberculosis H37Rv (Via et al., 1996) and to PhoP of Bacillus subtilis (Lee and Hulett, 1992). Based on this data, the DNA sequence (440-bp fragment, Fig. 3) and the derived polypeptide might play a role in regulation of virulence in mycobacteria.

1.4 Cloning, expression and purification of the ca. 15kDa protein fused to thioredoxin

The λgt11 clone C9-2 (Fig. 4) was used as template to amplify a PCR fragment of 951-bp (Fig. 2, sequence position 451-1378) including the ORF for the ca. 15kDa protein (Fig. 5) and cleavage sites for the restriction endonucleases SmaI and SalI at the 5'- and 3'-ends. Amplification of the SmaI-SalI mycobacterial DNA fragment for insertion into pTrxFus (Invitrogen, Leek, The Netherlands) was done using the oligonucleotide primers with the sequence 5'-TCTAGACATATGACGCGAGTGGGTGTGATTTCG-3' (PRIMER 7, forward) and 5'-CATATGGTCGACCTAGGGCGTGTCTCCCAA-3' (PRIMER 8, reverse) corresponding to sequence positions 451-474 and 1378-1361 (Fig. 2). Composition of the reaction mix was the same as described above with 400 ng phage DNA as template. The probe was amplified in 30 cycles consisting of the same conditions as described. Cleavage sites were introduced by appropriate primers. After digestion with both restriction endonucleases the product was inserted in pTrxFus (Invitrogen, Leek, The Netherlands) to form the plasmid pCH3-8.

The E. coli strain GI724 was electroporated with the plasmid pCH3-8. Bacterial cultures (200 ml of Induction Medium (Invitrogen, Leek, The Netherlands) supplemented with 100 µg/ml ampicillin) grown at 30°C were induced to synthesize the fusion protein by tryptophan addition (100µg/ml) and temperature shift to 37°C. Cells were collected after 4 hours (10 000 x g, 5 min, 4°C), resuspended in 4 ml Osmotic Shock Solution (Invitrogen, Leek, The Netherlands), broken by three rounds of alternate sonication on ice (10 sec.) and shock freezing in liquid nitrogen, and pelleted (10 000 x g, 15 min, 4°C). Most of the fusion protein accumulated in the form of inclusion bodies and only a small fraction was present as soluble protein inside the cells.The pellet containing the inclusion bodies was resuspended (denaturation) in 10 ml 6 M guanidine/HCl (pH 8.5), incubated for 2 hours at room temperature and pelleted again (10 000 x g, 30 min, 4°C). The recombinant fusion protein was refolded by dialysing against 50 mM Tris/HCl (pH 8.0). Anion exchange chromatography was done with the help of a BioCAD perfusion system (Perseptive Biosystems) on a Poros column HQ/M (Perseptive Biosystems).

Elution was performed using a linear NaCl gradient (0-1M). The fusion protein concentration was determined with the BioRad Protein Assay Kit (BioRad, Munich, Germany). Purity was assessed by densitometry (Molecular Dynamics, Software Image Quant) and analytical SDS-PAGE and coomassie staining.

The ca. 15kDa protein fused to thioredoxin was refolded as described above. Further purification of the ca. 15kDa protein fused to thioredoxin was carried out by anion exchange chromatography (Fig. 8, A lane 3 and B lane1). After enterokinase cleavage of the purified ca. 15kDa protein fused to thioredoxin two protein bands were detectable on SDS-PAGE (Fig. 8, lane 2). By western blotting with a thioredoxin monoclonal antibody the lower 11kDa band was identified to be thioredoxin. The upper band corresponds to the ca. 15kDa recombinant protein of M. tuberculosis. This is the first report of expression and purification of the ca. 15kDa protein of M. tuberculosis in E. coli.

1.4. **Species specific diagnosis of mycobacteria :**

Deprotected and desalted Oligonucleotide primers were obtained from Gibco BRL (Eggenstein, Germany) or Eurogentec (Seraing, Belgium).

The oligodeoxyribonucleotide primers with the sequence 5'-GTCCATGTGCCGCCG CTG-3' (PRIMER 3, forward) and 5'-CTGCGCGGCTCCCGGCA-3' (PRIMER 4, reverse), specific for the DNA regions of the 2253-bp M. tuberculosis chromosomal region shown in Fig. 2 were used in PCR experiments to amplify a 377-bp fragment.

For amplification of a 380-bp fragment from the 440-bp chromosomal fragment, the oligodeoxyribonucleotide primers with the sequences 5'-CGAGGCTGAACGGCT TTG-3' (PRIMER 5, forward) and 5'-TCAACGTCCGCGGCAAGC-3' (PRIMER 6, reverse) corresponding to the DNA region shown in Fig. 3 were used. Amplifications were performed in 0.2 ml Micro Amp Reaction Tubes (Perkin Elmer, Norwalk, Connecticut, USA) in a final volume of 100 µl using a GeneAmp® PCR Kit (Perkin Elmer, Branchburg, New Jersey, USA). Reaction mixtures contained 10 mM Tris/HCl (pH 8.3), 50 mM KCl, 3 mM $MgCl_2$, 200 µM dNTP, 0.1 µM Primer, 30-100 ng chromosomal DNA from mycobacterial cell extracts (Table 1) and 2.5 U AmpliTaq® DNA polymerase. All components of a PCR reaction except for the template are included in the Kit. The reactions were performed using the automated Thermal Cycler Gene Amp PCR System 9600 (Perkin Elmer, Norwalk, Connecticut, USA). The samples were amplified by 40 cycles consisting of denaturation at 96°C for 2 min, annealing of the

primers at 25°C for 1 min and primer extension at 72°C for 3 min.

After amplification, 10 µl of each product was electrophoresed in a horizontal 1.5% agarose gel. Gels were precasted using a 1:10 000 dilution of SYBR Green I stock reagent (Eugene, Leiden, The Netherlands) in 10 mM Tris/HCl, 1 mM EDTA (pH 8.0).

For DNA sequencing the appropriate 377-bp and 380-bp PCR products from the mycobacterial cell extract samples (Table 1) were purified from an 1.5% agarose gel using a Gel Extraction Kit (QIAGEN, Hilden, Germany).

1.4.1. **The 377-bp region:**

The 377-bp region (Fig.2) of the isolated and sequenced 2253-bp M. tuberculosis chromosomal fragment and the 380-bp region (Fig.3) of the identified 440-bp chromosomal fragment were examined for their suitability for strain differentiation (Table 1). A PCR-product of the predicted size and a 100% DNA sequence homology in the 377-bp region was detected only in the members of the M. tuberculosis complex. No amplification product was obtained from other mycobacteria (Table 1). Therefore, the PCR primers of the 377-bp region are useful for the rapid discrimination of M. tuberculosis complex (M. tuberculosis, Mycobacterium bovis, Mycobacterium bovis BCG, Mycobacterium africanum and Mycobacterium microti) from other mycobacteria.

1.3.2. The 380-bp region:

A predominant amplification product of correct size of the 380-bp region was obtained from the chromosomal DNA samples of the M. tuberculosis complex including the vaccine strain M. tuberculosis BCG, the tuberculosis isolate Tub118 and the mycobacterial species M. asiaticum, M. gastri, M. gordonae and M. kansasii. Thus, this fragment can be used for the identification of above mycobacterial species, since no amplification product was obtained from other mycobacterial species (Table 1).

2. Clone containing PGRS Element

2.1. Cloning of DNA fragment containing PGRS elements:

We screened Lawrist cosmid library of M. tuberculosis DNA using a degenerate oligonucleotide of the sequence 5'-C/GGCC/GGCC/GGGC/GACC/GGGC/GGGC/GGCCGGCTCC/GGG- 3' which was designed in such a way that it contained GC rich regions as well as it coded for a putative proline rich polypeptide. Colony hybridization using labelled oligonucleotide was performed using standard

procedures (Maniatis et al.1982). Filters were prehybridized and probed at 42°C overnight in a solution containing 6xSSC, 1 mM Sodium phosphate, 1mM EDTA, 0.05% skimmed milk, 0.5%SDS. Filters were washed twice in 2xSSC;0,3%SDS for 15 min at 65°C.
First screeinig yielded six positive clones which were recheked by hybridization with the oligonucleotide. Three clones gave strong signal and restriction mapping of the clones showed identical restriction pattern. Further restriction mapping and Southern hybridization of one of the clones called identified an about 4kb SalI fragment that hybridized strongly to the oligonucleotide.

2.2. DNA sequencing of the cloned fragment: The ca. 4kb SalI fragment was subcloned in pUC19 and the clone was named Mtub-Clara-Klon. Entire insert was sequenced by primer walking method. The DNA sequence is presented in figure 9. There were unusual difficulties in obtaining the sequence of the recobminant clone because of the high GC rich content and due to the presence of unusual repeats.

2.3. Proteins coded by the cloned DNA:

We identified at least 4 ORF (open reading frames) belonging to a ca. 9kDa, 55kDa, 74kDa and a 77kDa protein (Fig. 10). Interestingly, the amino acid sequence of the 9kDa, 55kDa ,74kDa and the 77kDa proteins didnot show strong homology to any sequences reported so far for Mycobacteria (Genbank and Swissprot Databases) . In addition, the 9kDa, 55kDa and the 74kDa proteins have an unusually high content proline, nevertheless, no strong homology with the known proline rich antigens (Laqueyrerie et al. 1995; Infect.Immun.63.4003) of mycobacteria was observed. Unexpectedly, the amino acid sequence showed restricted homology to Mucein like proteins from eucaryotes. The 77kDa protein is highly rich in amino acid glycine and may be a cell wall protein of Mycobacterium tuberculosis. Such proteins have not been reported from M. tuberculosis.

2.3. DNA finger-printing:

The ca. 4kb SalI fragment was used to probe (Southern hybridization) genomic DNA of different mycobacteria digested by PvuII (Fig. 11). The results show that each strain showed a characteristic pattern making the differentiation of M. tuberculosis-Rv, M. tuberculosis-Ra, M. bovis and the M. tuberculosis Erdman strain. The ca. 4kb SalI fragment is also suitable for finger printing of clinical isolates, since hybridization of the probe to the genomic DNA of clinical isolates from tuberculosis patients also yielded strain specific finger print (Fig. 12). No hybridization to the genomic DNA of M. smegmatis, M. vaccae, M. avium, M. chelonie, M.

fortituim, M. phlei was observed.

REFERENCES

Aminosante, M., Bamini, S., Corsini, V., Paone, G., Read Jr, C.A., Tartoni, P.L., Singh, M., Albera, C., Bisetti, A., Senesi, S., Campa, M. and Saltini, C.: Evaluation of a novel tuberculosis complex-specific 34 kDa protien in the serological diagnosis of tuberculosis. Er. Respir. J. **8** (1995) 2008-2014

Barnes, P.F., Bloch, A.B., Davidson, P.T. and Snider Jr., D.E.: Tuberculosis in patients with human immunodeficiency virus infection. New Engl. J. Med. **324** (1991) 1644-1650

de Wit, D., Maartens, G. and Steyn, L.: A comparative study of the polymerase chain reaction and conventional procedures for the diagnosis of tuberculous pleural effusion. Tuber. Lung Dis. **73** (1992) 262-267

Glennon, M., Smith, T., Cormican, M., Noone, D., Barry, T., Maher, M., Dawson, M., Gilmartin, J.J. and Gannon, F.: The ribosomal intergenic spacer region: a target for the PCR based diagnosis of tuberculosis. Tuber. Lung Dis. **75** (1994) 353-360

Grybowski, S.: Tuberculosis in the third world. Thorax **46** (1991) 689-691

Huynh, T.V., Young, R.A. and Davis, R.W.: In: DNA Cloning., Glover, D.M., ed., Vol. 1, IRL, Oxford, England (1985), 56-110

Kochi, A.: The global tuberculosis situation and the new control strategy of the Word Health Organization. Tubercle **72** (1991) 1-6

Lee, J.W. and Hulett, F.M.: Nucleotide sequence of the phoP gene encoding PhoP, the response regulator of the phosphate regulon of Bacillus subtilis. Nucleic Acids Res. **20** (1992) 5848

Maniatis, T., Fritsch, E.F. and Sambrook: Molecular cloning: a laboratory manual. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mariani, F., Piccolilla, E., Colizzi, V., Rappuoli, R. and Gross, R.: Characterization of an IS-like element from Mycobacterium tuberculosis. J. Gen. Microbiol. **139** (1993), 1767-1772

Miller, S.I.:PhoP/PhoQ: macrophage-specific modulators of Salmonella virulence? Mol. Microbiol. **5** (1991), 2073-2078

Miller, J.F., Mekalanos, J.J. and Falkow, S.: Coordinate regulation and sensory transduction in the control of bacterial virulence. Science **243** (1989), 916-922

Narita, M.: Polymerase chain reaction for diagnosis of infectious disease. Acta Paediatr.

Jpn. 35 (1993), 89-97

Pearson, W.R., Lipman, D.J.: Improved tools for biological sequence comparison. Proc. Nat. Acad. Sci. 85 (1988), 2444-2448

Selwyn, A.P., Hartel, D., Lewis, V.A., Schoenbaum, E.E., Vermund, S.H., Klein, R.S., Walker, A.T. and Friedland, G.H.: A prospective study of the risk of tuberculosis among intravenous drug users with the human immunodeficiency virus infection. New Engl. J. Med. 320 (1989) 545-550

Shankar, P., Manjunath. N. and Mohan, K.K.: Rapid diagnosis of tuberculous meningitis by polymerase chain reaction. Lancet, 337 (1991), 5-7.

Via, L.E., Curcic, R., Mudd, M. H., Dhandayuthapani, S., Ulmer, R.J., and Deretic,V.: Elements of signal transduction in Mycobacterium tuberculosis: in vitro phosphorylation and in vivo expression of the response regulator MtrA. J. Bacteriol. 178 (1996) 3314-3321

Vismara, D., Filippone Mezzopreti, M., Gilardini, M., Del Porto, P., Lombardi, G., Piccolella, E., Damiani, G., Rappuoli, R. and Colizzi, V.: Identification of a 35-Kilodalton Mycobacterium tuberculosis protein containing B- and T-cell epitopes. Infect. Immun. 58 (1990) 245-251

Wilkins, E.G.L.: The serodiagnosis of tuberculosis. In: Clinical Tuberculosis. Davies P.D.O. (ed.) (1994) Chapman & Hall, London.

Young, R.A., Bloom, R.B., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis R.W.: Dissection of Mycobacterium tuberculosis antigens using recombinant DNA. Proc. Natl. Acad. Sci. USA 82 (1985) 2583-2587.

Young, R.A. and Davis, R.W. (1985) Constructing and screening cDNA library in λgt10 and λgt11, p. 49-78. In: Genetic engineering: principles and methods, J.K. Setlow and A. Hollaender (ed.), vol. 7. Plenum Publishing Corp., New York.

Table 1 Distribution of the 377-bp sequence and the 380-bp sequence in different mycobacteria

	Species	Presence of the 377-bp fragment	Presence of the 380-bp fragment
M. tuberculosis **complex**	*M. tuberculosis* H37Rv	+	+
	M. tuberculosis H37Ra	+	+
	M. bovis	+	+
	M. africanum	+	+
	M. microti	+	+
vaccine strain	*M. bovis BCG*	+	+

	Species	Presence of the 377-bp fragment	Presence of the 380-bp fragment
clinical tuberculosis isolate	Tub 118	+	+
	M. asiaticum	-	+
	M. avium	-	-
	M. chelonae	-	-
	M. flavescens	-	-
	M. fortuitum	-	-
	M. parafortuitum	-	-
	M. gastri	-	+
	M. gordonae	-	+
	M. intracellulare	-	-
	M. kansasii	-	+
	M. lufu	-	-
	M. marinum	-	-
	M. nonchromogenium	-	-

	Species	Presence of the 377-bp fragment	Presence of the 380-bp fragment
	M. pergrinum	-	-
	M. phlei	-	-
	M. scrofulaceum	-	-
	M. simiae	-	-
	M. smegmatis	-	-
	M.terrae	-	
	M. ulcerus	-	
	M. vaccae	-	
	M. xenopi	-	
	M. thermoresistibile	-	-
	M. triviale	-	-
none tuberculosis strains	*Nocardia asteroides*	-	-
	Rodococcus equi	-	-

Claims

1. (I) DNA
(a) having sequence (I) according to **figure 9**, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),
(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e); or

(II) DNA
(a) having sequence (VI) according to **figure 2**, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),
(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e); or

(III) DNA
(a) having sequence (IX) according to **figure 3**, wherein optionally one or more condons can be replaced by condons coding for the same amino acid(s),

(b) having a sequence complementary to said of (a),
(c) being single-stranded, wherein its strand is hybridizable with that of the DNA according to (a) or (b),
(d) being double-stranded, the sequences of its single strands being defined as in (a) and (b), respectively,
(e) being double-stranded, its single strands being hybridizable with those of the DNA according to (d), or
(f) being a subsequence of the sequences according to (a) to (e).

2. A DNA according to claim 1 (I)(c), (I)(e), (II)(c), (II)(e), (III)(c) or (III)(e), its single strands being hybridizable at a temperature of at least 25 ^{o}C and at a concentration of NaCl of 1 M.

3. RNA being a transcript of a DNA according to claim 1 or 2.

4. Protein being encoded by a DNA according to claim 1 or 2.

5. Protein having the amino acid sequence (II) according to figure 13.

6. An about 74 kDa protein according to claim 4 or 5.

7. Protein having the amino acid sequence (III) according to figure 14.

8. An about 77 kDa protein according to claim 4 or 7.

9. Protein having the amino acid sequence (IV) according to figure 15.

10. An about 9 kDa protein according to claim 4 or 9.

11. Protein having the amino acid sequence (V) according to figure 16.

12. An about 55 kDa protein according to claim 4 or 11.

13. Protein having the amino acid sequence (VII) according to figure 5.

14. An about 15 kDa protein according to claim 4 or 13.

15. Protein having the amino acid sequence (VIII) according to figure 6.

16. An about 31 kDa protein according to claim 4 or 15.

17. Protein having the amino acid sequence (X) according to figure 7.

18. An about 17 kDa protein according to claim 4 or 17.

19. A protein according to any of claims 4 to 18, wherein the protein is a recombinant protein, especially a protein produced by means of a bacterial strain, a yeast strain, a fungal strain or a cell line of a higher eucaryote.

20. A protein being encoded by a DNA sequence according to claim 1 or 2 and which can be recovered by a method comprising the following steps:
(i) subjecting proteins encoded by said DNA sequence to a usual test for diagnosis of tuberculosis,
(ii) selecting a protein showing an inhibitory effect and
(iii) isolating and recovering said protein.

21. DNA according to claim 1 or 2, RNA according to claim 3 or protein according to any of claims 4 to 20 which can be used for

(i) diagnosis of tuberculosis in humans and animals and/or
(ii) diagnosis of other mycobacterial infections in humans or animals,
each especially by means of samples taken from humans or animals.

22. Use of a DNA according to any of claims 1, 2 or 21 for the identification of mycobacteria in media samples.

23. Use according to claim 22, comprising the steps of
(i) isolating the mycobacterium,
(ii) preparing crude or purified genomic DNA,
(iii) hybridizing it to a DNA according to claim 1 or 2 and
(iv) detecting the fragment pattern using conventional methods such as a radioactivity assay, chemiluminiscence or fluorescence.

24. Use according to any of claims 22 to 23, wherein as samples clinical samples are used.

25. Use of a DNA according to any of claims 1 to 2 or of a protein according to any of claims 4 to 20 for
(i) epidemeological purposes and/or
(ii) vaccination follow-up
for humans or animals suffering from mycobacterial infections, especially tuberculosis.

26. Use of a DNA according to any of claims 1 to 2 or of a protein according to any of claims 4 to 20 for the development of drugs useful for combating mycobacterial infections of humans or animals, especially tuberculosis, especially for testing and recovering of substances inhibiting mycobacterial infections in humans and animals, especially tuberculosis.

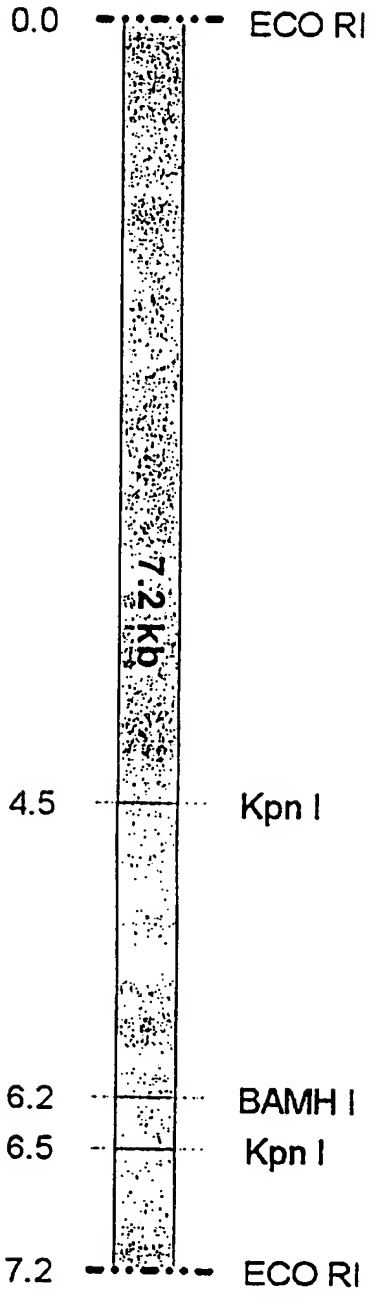


Fig. 1 Restriction endonuclease map of the 7.2 kb *M. tuberculosis* chromosomal region

Fig. 2 - 1/4

```
1   CGACGGCCGT GTGATCGGCC TCGTGGTCGA GGACTCTGCG GAAGCACCGT CGACGGGCAC CGGACCGAAC GCGGCACCCT TCTATCGCGG CATCCCGTCG
    GCTGCCGGCA CACTAGCCGG AGCACCAGCT CCTGAGACGC CTTCGTGGCA GCTGCCCGTG GCCTGGCTTG CGCCGTGGGA AGATAGCGCC GTAGGGCAGC

101 AGCGAAGTGA TCCGAGCACT CGACGAACTC GACTTCGGCG GCATCGTGGA GATGGACACC TTCCCATAGG CACAACCCGG GACGTATCAA CACGCACCAA
    TCGCTTCACT AGGCTCGTGA GCTGCTTGAG CTGAAGCCGC CGTAGCACCT CTACCTGTGG AAGGGTATCC GTGTTGGGCC CTGCATAGTT GTGCGTGGTT

201 ACCCTTCAAT CCGGTGGTAG TGGGTCAGTT TGATTTTCGC CGAACTCGAC GCGGCGTAGG CTCGCTGTGT CTAGGCAGGG CATGGTGTGG CTAGGCAAGG
    TGGGAAGTTA GGCCACCATC ACCCAGTCAA ACTAAAAGCG GCTTGAGCTG CGCCGCATCC GAGCGACACA GATCCGTCCC GTACCACACC GATCCGTTCC

301 CTAGGAATGG CTTGGCGTGG CAAGGCAAGG TTTGAGCCAG CGGGCCGCAG CCATAGGAGA TGCGCCGAAC AGGCCAAAGT ACTACCGCTG CAGCCCGCGC
    GATCCTTACC GAACCGCACC GTTCCGTTCC AAACTCGGTC GCCCGGCGTC GGTATCCTCT ACGCGGCTTG TCCGGTTTCA TGATGGCGAC GTCGGGCGCG

                                                ORF 1 →  ValThrArgVal GlyValIle SerAspGlu PheTrpAlaVal ValGluPr
                                                         PRIMER 1                                    PRIMER 2
401 CTAGGGCGTG TCTCCCAAAT TTTTAGGTAC TGGCCAGCGA GGATTGGCCG GTGACGCGAG TGGGTGTGAT TTCGGACGAG TTCTGGGCCG TGGTCGAGCC
    GATCCCGCAC AGAGGGTTTA AAAATCCATG ACCGGTCGCT CCTAACCGGC CACTGCGCTC ACCCACACTA AAGCCTGCTC AAGACCCGGC ACCAGCTCGG

        ORF 2 ➔ MetArg AlaSerPro AlaAspGly LeuAlaIleThr GlyLeuSer TrpLysGly SerArgGlyGly SerValArg GluValArgGl

ORF 1  oLeuMetPro SerHisGluGly LysProGly ArgArgPhe SerAspHis ArgLeuIleLeu GluGlyIle AlaTrpArgPhe ArgThrGly SerPr Trp
501 GTTGATGCCG TCGCATGAGG GCAAGCCCGG CAGACGGTTT AGCGATCACC GGCTTATCCT GGAAGGGATC GCGTGGCGGT TCCGTACGGG AAGTCCGTGG
    CAACTACGGC AGCGTACTCC CGTTCGGGCC GTCTGCCAAA TCGCTAGTGG CCGAATAGGA CCTTCCCTAG CGCACCGCCA AGGCATGCCC TTCAGGCACC

ORF 2  yGlyThrCys ProLeuSer SerGlyArg GlyLysArgCys GlySerAla IleThrValGly ArgTrpMet ValProAla ThrArgCys SerProThrLeu

ORF 1  ArgAspLeuPro AlaGluPhe GlyProTrp GlnThrValTrp LysArgHis HisArgTrp SerLeuAspGly ThrCysAsp GluValPhe AlaHisValA

                                                                                    KpnI
                                                                                    ~~~~~~
601 CGGGACCTGC CCGCTGAGTT CGGGCCGTGG CAAACGGTGT GGAAGCGCCA TCACCGTTGG TCGCTGGATG GTACCTGCGA CGAGGTGTTC GCCCACGTTG
    GCCCTGGACG GGCGACTCAA GCCCGGCACC GTTTGCCACA CCTTCGCGGT AGTGGCAACC AGCGACCTAC CATGGACGCT GCTCCACAAG CGGGTGCAAC
```

Fig. 2 - 2/4

```
ORF 2  ProArgCys SerGlyTrp ThrLeuArgTrp ProArgIle SerArgSerCys CysArgTrp IleProArg ThrCysGly HisThrSerIle ArgArgAlaP

ORF 1  laAlaValPhe GlyValAsp AlaGluValAla GluAspIle GluLysLeu LeuSerValAsp SerThrAsn ValArgAla HisGlnHisSer AlaGlyAl
  701  CCGCGGTGTT CGGGGTGGAC GCTGAGGTGG CCGAGGATAT CGAGAAGCTG CTGTCGGTGG ATTCCACGAA CGTGCGGGCA CACCAGCATT CGGCGGGCGC
       GGCGCCACAA GCCCCACCTG CGACTCCACC GGCTCCTATA GCTCTTCGAC GACAGCCACC TAAGGTGCTT GCACGCCCGT GTGGTCGTAA GCCGCCCGCG

ORF 2  roAlaArg ThrArgSerPro GlnGlyAla LeuSerAsp TyrLysLysSer AlaAspGlu ProAspAsp HisAlaIleGly ArgSerArg GlyGlyLeuTh

ORF 1  aCysSerAsp ThrLeuAlaThr GlyGlyThr ValGlyLeu GlnGluIleA rgArg*** ←
  801  CTGCTCGGAC ACGCTCGCCA CAGGGGGCAC TGTCGGATTA CAAGAAATCC GCCGATGAAC CCGACGATCA TGCGATCGGC CGCTCGCGCG GCGGGCTGAC
       GACGAGCCTG TGCGAGCGGT GTCCCCCGTG ACAGCCTAAT GTTCTTTAGG CGGCTACTTG GGCTGCTAGT ACGCTAGCCG GCGAGCGCGC CGCCCGACTG

ORF 2  rThrLysIle HisAlaLeu ThrAspGln ArgGluAlaPro ValArgIle ArgLeuThr AlaGlyGlnAla GlyAspAsn ProGlnLeu LeuProLeuLeu
                                                  BamHI
                                                  ~~~~~~
  901  CACCAAGATC CATGCCCTGA CCGATCAGCG CGAAGCCCCG GTGCGGATCC GGTTGACCGC AGGCCAGGCC GGCGACAACC CGCAACTGCT GCCCCTGCTC
       GTGGTTCTAG GTACGGGACT GGCTAGTCGC GCTTCGGGGC CACGCCTAGG CCAACTGGCG TCCGGTCCGG CCGCTGTTGG GCGTTGACGA CGGGGACGAG

ORF 2  AspAspTyr ArgHisAla IleThrGluTyr AlaLeuGly SerThrAspPhe ArgLeuLeu AlaLeuLys AlaTyrSer HisProSerThr ArgAlaAlaL
 1001  GACGACTATC GCCATGCCAT CACCGAATAC GCCCTGGGCA GCACGGATTT CCGCTTACTC GCCGACAAGG CCTACTCACA CCCAAGTACC CGTGCCGCAT
       CTGCTGATAG CGGTACGGTA GTGGCTTATG CGGGACCCGT CGTGCCTAAA GGCGAATGAG CGGCTGTTCC GGATGAGTGT GGGTTCATGG GCACGGCGTA

ORF 2  euArgSer LysLysIleLys HisThrIle ProGluArg GlnAspGlnIle AspArgArg LysAlaLys GlySerAlaGly GlyArgPro ProAlaPheAs
 1101  TACGGTCTAA GAAGATCAAG CACACCATCC CCGAACGCCA AGATCAGATC GACCGGCGCA AGGCCAAGGG GTCTGCCGGC GGGCGGCCAC CAGCATTCGA
       ATGCCAGATT CTTCTAGTTC GTGTGGTAGG GGCTTGCGGT TCTAGTCTAG CTGGCCGCGT TCCGGTTCCC CAGACGGCCG CCCGCCGGTG GTCGTAAGCT

ORF 2  pAlaAlaLeu TyrGlyLeu ArgAsnThr ValGluArgGly PheHisArgLeu LysGln TrpArgGlyIle AlaThrArg TyrAspLys TyrAlaLeuThr
 1201  CGCCGCGCTC TACGGGCTAC GCAACACCGT CGAACGCGGC TTCCATCGAC TCAAGCAGTG GCGCGGCATC GCAACCCGCT ACGACAAATA CGCCCTGACC
       GCGGCGCGAG ATGCCCGATG CGTTGTGGCA GCTTGCGCCG AAGGTAGCTG AGTTCGTCAC CGCGCCGTAG CGTTGGGCGA TGCTGTTTAT GCGGGACTGG

ORF 2  TyrLeuGly GlyValLeu LeuAlaCysAla ValIleHis AlaArgValGlyThrProLys LeuGlyAsp ThrPro*** ←
 1301  TACCTCGGCG GCGTCCTGCT GGCCTGCGCC GTCATCCACG CCCGAGTGGG AACTCCGAAA TTGGGAGACA CGCCCTAGTC GGGATACCTG CGGCACCGGT
       ATGGAGCCGC CGCAGGACGA CCGGACGCGG CAGTAGGTGC GGGCTCACCC TTGAGGCTTT AACCCTCTGT GCGGGATCAG CCCTATGGAC GCCGTGGCCA
```

Fig. 2 - 3/4

```
1401  CCAGCTGGCT GTCCGTGACG GCGTGCGCGG TGCGCTGGTC AAGGACGTGC GGTTGCCGCA AGCATTGATG CCCGAGGTTG TGCTCAACTG CGACCGGGAC
      GGTCGACCGA CAGGCACTGC CGCACGCGCC ACGCGACCAG TTCCTGCACG CCAACGGCGT TCGTAACTAC GGGCTCCAAC ACGAGTTGAC GCTGGCCCTG

1501  GATCCCGGGC AACGTGGGCA CAACGTCGAG GCGATCCAGC AGGCGCTTGC TCGGTTCGCA GCACCTGCCG ACCAAGCGCT CCTACCCACG TTCCGCGCTT
      CTAGGGCCCG TTGCACCCGT GTTGCAGCTC CGCTAGGTCG TCCGCGAACG AGCCAAGCGT CGTGGACGGC TGGTTCGCGA GGATGGGTGC AAGGCGCGAA

                                                                                  PRIMER 3
1601  TCGACCTACC TCGTGTTCGA CGCATTGATC GCCCACGGTG ATCGTCACGA CCGCAACTGG GCGGTCCATG TGCCGCCGCT GGAGACAAAG TATGTGGACG
      AGCTGGATGG AGCACAAGCT GCGTAACTAG CGGGTGCCAC TAGCAGTGCT GGCGTTGACC CGCCAGGTAC ACGGCGGCGA CCTCTGTTTC ATACACCTGC

1701  CGCTGTGCCC ATCGTTCGAC CACGCGGCCA GCCTCGGGTT CACGCTGACC GACCAGACCC CCGCTCAGCA TTTGCACGAC GGGTGAGCTT CCCCTCTTGT
      GCGACACGGG TAGCAAGCTG GTGCGCCGGT CGGAGCCCAA GTGCGACTGG CTGGTCTGGG GGCGAGTCGT AAACGTGCTG CCCACTCGAA GGGGAGAACA

1801  TGTGGAGACA TCGCGTAGCC GCGCCAACTC GCTGGCGAGC TTTTCTGCCA CGTCTTCGCC GAGTAAACGA TCGCGCTCCG GTCGCGCCGG CTTGAAGAGC
      ACACCTCTGT AGCGCATCGG CGCGGTTGAG CGACCGCTCG AAAAGACGGT GCAGAAGCGG CTCATTTGCT AGCGCGAGGC CAGCGCGGCC GAACTTCTCG

1901  CACGTCCGCT TCTGCGAAGA ATGTTTGAGC CAATCGTGCG GGTGTTGCCC CTGGGTCTCC AGGCCGGCAA CAGACCAGTC GTCAACGCAA ACGACCGGCC
      GTGCAGGCGA AGACGCTTCT TACAAACTCG GTTAGCACGC CCACAACGGG GACCCAGAGG TCCGGCCGTT GTCTGGTCAG CAGTTGCGTT TGCTGGCCGG

2001  AGGAATCGGG CACGAACTCA TTGTGCCGGG AGCCGCGCAG ACGGCGTATC GATCAGCGCG CCAACGCCCC CATCTGGATG CCGACAGTGC TGGACAGTGG
      TCCTTAGCCC GTGCTTGAGT AACACGGCCC TCGGCGCGTC TGCCGCATAG CTAGTCGCGC GGTTGCGGGG GTAGACCTAC GGCTGTCACG ACCTGTCACC
                              PRIMER 4

2101  CTTCGACCCG GCATATTGGC AATTAGAAAC GACTCTTTCG AAGATTACCG GCACTAGCAA ACGCCAGCAC CCGTGAGCGC GGATTGTCCG CTGCCTCTGC
      GAAGCTGGGC CGTATAACCG TTAATCTTTG CTGAGAAAGC TTCTAATGGC CGTGATCGTT TGCGGTCGTG GGCACTCGCG CCTAACAGGC GACGGAGACG

2201  CAATCGGGCT GTGGAGGCTT GACGGCGAAT GTAGCGCTTG GACCCGATCC AGTGCACCAA CCCTACCGCG ACACGCCAGG TGCGAAAGAG CCGTCATTCG
      GTTAGCCCGA CACCTCCGAA CTGCCGCTTA CATCGCGAAC CTGGGCTAGG TCACGTGGTT GGGATGGCGC TGTGCGGTCC ACGCTTTCTC GGCAGTAAGC

2301  CCTAGAATCC GCGCGCCGCG CGTCTCATTT GTGTGCCGAA GTGTGCTAGC CCGGGCACCA GATCACACCG CCATGTCAGC CGCCAAGACC TTGTTAAGGA
      GGATCTTAGG CGCGCGGCGC GCAGAGTAAA CACACGGCTT CACACGATCG GGCCCGTGGT CTAGTGTGGC GGTACAGTCG GCGGTTCTGG AACAATTCCT

2401  TGCTGACGTA AACGTTAAAA AACCGTAAAC GTCGGGCGTC CCATGCCTTT CGGTCGCGAC TGTGGAGGCT TGGCCGGTTA GCGTTACGCC AATCCACAGG
      ACGACTGCAT TTGCAATTTT TTGGCATTTG CAGCCCGCAG GGTACGGAAA GCCAGCGCTG ACACCTCCGA ACCGGCCAAT CGCAATGCGG TTAGGTGTCC
```

```
2501  CAGCCGTACG CAAACACACA GGGTCGGCCC GGGGTGCTAG CGGCCGGCAG GTTTCACACA CAAGGGAGCA ATGGATGTCA TTCCTCAAGA CAGTGCCTGA
      GTCGGCATGC GTTTGTGTGT CCCAGCCGGG CCCCACGATC GCCGGCCGTC CAAAGTGTGT GTTCCCTCGT TACCTACAGT AAGGAGTTCT GTCACGGACT

2601  AGAGTTGACG GCTGCGGCCG CGCAACTCGG GACCATCGGT GCTGCGATGG CAG
      TCTCAACTGC CGACGCCGGC GCGTTGAGCC CTGGTAGCCA CGACGCTACC GTC
```


Fig. 2 2253-bp *M. tuberculosis* chromosomal region including BamHI, EcoRI and KpnI restriction sites and oligonucleotides for screening the λgt11 *M. tuberculosis* library (PRIMER 1 and PRIMER 2 underlined) and for amplification of the 377-bp region (377-bp region in bold, PRIMER 3 and PRIMER 4 underlined). Amino acid sequences of the ca. 15kDa and the ca. 31kDa proteins are shown above the DNA sequences and are marked with arrows (→ ca. 15kDa ORF 1, ➔ ca. 31kDa ORF 2).

PRIMER 5

1 CGAGGCTGAA CGGCTTTGTC ATGTAGTCGT CGGCGCCCGA GGTCAGACCG GTGACCCGGT CCATCACGGA ATCGCCGCGG GTGAGGAACA GCGTGGGTGT
GCTCCGACTT GCCGAAACAG TACATCAGCA GCCGCGGGCT CCAGTCTGGC CACTGGGCCA CGTAGTGCCT TAGCGCGCGC CACTCCTTGT CGCACCCACA

→ LeuSerPhe ProLysThrMet TyrAspAsp AlaGlySer ThrLeuGlyThr ValArg AspMetValSer AspArgAlaThr LeuPheLeu ThrProThr

101 GTAGACGTCG GATTCTCGGA CCCGTCGCAG GATTTCCAAC CCGTCCACAT CGGGAAGCAT GATGTCGAGG ACCAGCACAT CGCGGCCGAC CTTGTCGAAC
CATCTGCAGC CTAAGAGCCT GGGCAGCGTC CTAAAGGTTG GGCAGGTGTA GCCCTTCGTA CTACAGCTCC TGGTCGTGTA GCCCCGGCTG GAACAGCTTG

TyrValAsp SerGluArgVal ArgArgLeu IleGluLeuGly AspValAsp ProLeuMet IleAspLeuVal LeuValAsp ProGlyVal LysAspPheL

201 TTGGCTATGG CCTCTTGCCC GTCGTGGGCG ACTTCGACAT CCCAGCCTTC GTAGTGCAGC GCCATCTTGA CCAGATTGGT CAGCGCTGGT TCGTCATCGA
AACCGATACC GGAGAACGGG CAGCACCCGC TGAAGCTGTA GGGTCGGAAG CATCACGTCG CGGTAGAACT GGTCTAACCA GTCGCGACCA AGCAGTAGCT

ysAlaIleAla GluGlnGly AspHisAlaVal GluValAsp TrpGlyGlu TyrHisLeuAla MetLysVal LeuAsnThr LeuAlaProGlu AspAspVa

301 CCAACAACAC CCGGATCGGT GATCCATCCG CGCGATGAAT CCGTGGCAGC TGCCCCAAGA TGGCTTGCCG CGGACGTTGA CTGCGCGTGT ACCCCGACAT
GGTTGTTGTG GGCCTAGCCA CTAGGTAGGC GCGCTACTTA GGCACCGTCG ACGGGGTTCT ACCGAACGGC GCCTGCAACT GACGCGCACA TGGGGCTGTA

PRIMER 6

lLeuLeuVal ArgIleProSer GlyAspAla ArgHisIle ArgProLeuGln GlyLeuPhe AlaGlnArg ProArgGlnSer ArgThrTyr GlySerMet

401 CGTCGTCATG CTCCCGTATC CTCTCAATTC TGTGCAAGCG
GCAGCAGTAC GAGGGCATAG GAGAGTTAAG ACACGTTCGC

ThrThrMetSer GlyTyrGly ArgLeuGlu ThrCysAla ← ORF 3

Fig. 3 DNA sequence of the 440-bp *M. tuberculosis* chromosomal region including the 380-bp region (in bold) used in PCR experiments and the amino acid sequence of the ORF 3 shown below the complementary DNA strand (← ORF 3).

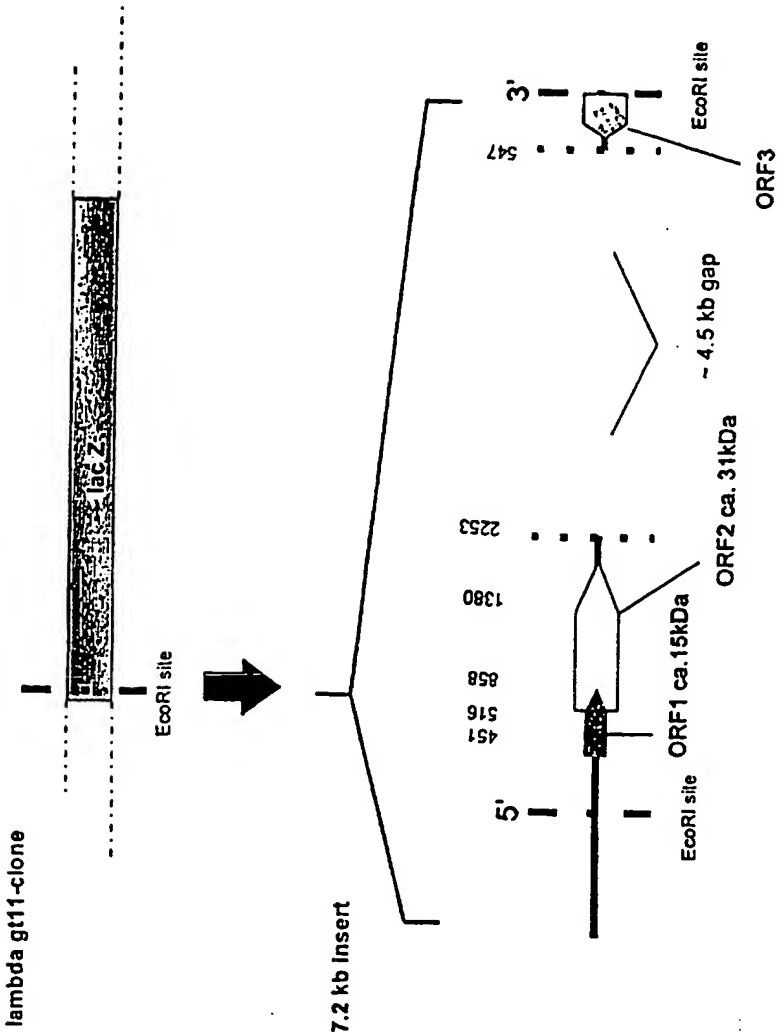


Fig. 4 Overview of the isolated λgt11-clone C9-2: 7.2 kb insert fragment, sequenced chromosomal regions (▬) and ORF1, ORF2 and ORF3 marked with arrows.

Val Thr Arg Val Gly Val Ile Ser Asp Glu Phe Trp Ala Val Val Glu Pro Leu Met Pro Ser His Glu Gly
Lys Pro Gly Arg Arg Phe Ser Asp His Arg Leu Ile Leu Glu Gly Ile Ala Trp Arg Phe Arg Thr Gly Ser
Pro Trp Arg Asp Leu Pro Ala Glu Phe Gly Pro Trp Gln Thr Val Trp Lys Arg His His Arg Trp Ser Leu
Asp Gly Thr Cys Asp Glu Val Phe Ala His Val Ala Ala Val Phe Gly Val Asp Ala Glu Val Ala Glu Asp
Ile Glu Lys Leu Leu Ser Val Asp Ser Thr Asn Val Arg Ala His Gln His Ser Ala Gly Ala Cys Ser Asp
Thr Leu Ala Thr Gly Gly Thr Val Gly Leu Gln Glu Ile Arg Arg ***

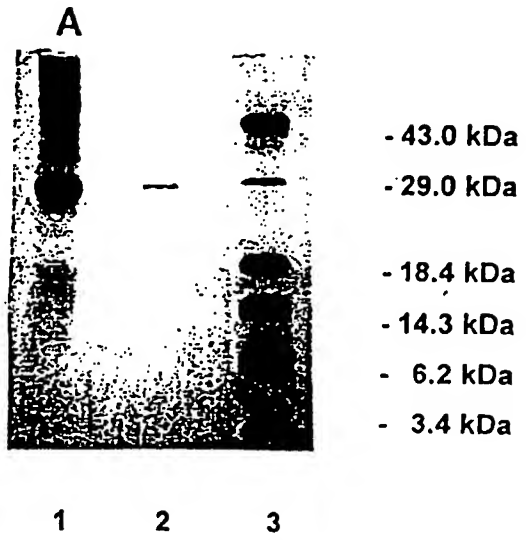
Fig. 5 Amino acid sequence of the ca. 15kDa protein (ORF 1)

Met Arg Ala Ser Pro Ala Asp Gly Leu Ala Ile Thr Gly Leu Ser Trp Lys Gly Ser Arg Gly Gly Ser Val Arg Glu Val
Arg Gly Gly Thr Cys Pro Leu Ser Ser Gly Arg Gly Lys Arg Cys Gly Ser Ala Ile Thr Val Gly Arg Trp Met Val Pro
Ala Thr Arg Cys Ser Pro Thr Leu Pro Arg Cys Ser Gly Trp Thr Leu Arg Trp Pro Arg Ile Ser Arg Ser Cys Cys Arg
Trp Ile Pro Arg Thr Cys Gly His Thr Ser Ile Arg Arg Ala Pro Ala Arg Thr Arg Ser Pro Gln Gly Ala Leu Ser Asp
Tyr Lys Lys Ser Ala Asp Glu Pro Asp Asp His Ala Ile Gly Arg Ser Arg Gly Gly Leu Thr Thr Lys Ile His Ala Leu
Thr Asp Gln Arg Glu Ala Pro Val Arg Ile Arg Leu Thr Ala Gly Gln Ala Gly Asp Asn Pro Gln Leu Leu Pro Leu Leu
Asp Asp Tyr Arg His Ala Ile Thr Glu Tyr Ala Leu Gly Ser Thr Asp Phe Arg Leu Leu Ala Leu Lys Ala Tyr Ser His
Pro Ser Thr Arg Ala Ala Leu Arg Ser Lys Lys Ile Lys His Thr Ile Pro Glu Arg Gln Asp Gln Ile Asp Arg Arg Lys
Ala Lys Gly Ser Ala Gly Gly Arg Pro Pro Ala Phe Asp Ala Ala Leu Tyr Gly Leu Arg Asn Thr Val Glu Arg Gly Phe
His Arg Leu Lys Gln Trp Arg Gly Ile Ala Thr Arg Tyr Asp Lys Tyr Ala Leu Thr Tyr Leu Gly Gly Val Leu Leu Ala
Cys Ala Val Ile His Ala Arg Val Gly Thr Pro Lys Leu Gly Asp Thr Pro ***

Fig. 6 Amino acid sequence of the ca. 31kDa protein (ORF 2)

Ala Cys Thr Glu Leu Arg Gly Tyr Gly Ser Met Thr Thr Met Ser Gly Tyr Thr Arg Ser Gln Arg Pro Arg
Gln Ala Phe Leu Gly Gln Leu Pro Arg Ile His Arg Ala Asp Gly Ser Pro Ile Arg Val Leu Leu Val Asp
Asp Glu Pro Ala Leu Thr Asn Leu Val Lys Met Ala Leu His Tyr Glu Gly Trp Asp Val Glu Val Ala His
Asp Gly Gln Glu Ala Ile Ala Lys Phe Asp Lys Val Gly Pro Asp Val Leu Val Leu Asp Ile Met Leu Pro
Asp Val Asp Gly Leu Glu Ile Leu Arg Arg Val Arg Glu Ser Asp Val Tyr Thr Pro Thr Leu Phe Leu Thr
Ala Arg Asp Ser Val Met Asp Arg Val Thr Gly Leu Thr Ser Gly Ala Asp Asp Tyr Met Thr Lys Pro Phe
Ser Leu

Fig. 7 Amino acid sequence of the ca. 17kDa protein



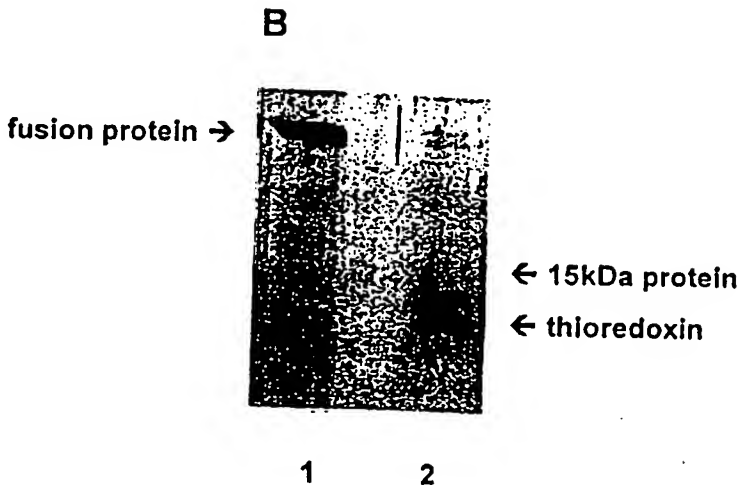


Fig. 8 A: SDS-PAGE of the insoluable pellet fraction (lane1) and the purified ca.15kDa recombinant antigen (lane2). Lane A3 shows protein molecular weight standards (2.850-43.000 molecular weight range; GIBCO BRL). B: SDS-PAGE of the purified ca. 15kDa thioredoxin fusion protein (lane1) and the two protein bands obtained after enterokinase cleavag (lane2).

Fig. 9
1/5

```
          10         20         30         40         50
GTCGACGTCT ACCGCACCTT CGTCGGCGAG ATGGACGACG AAGAGGCCGA

          60         70         80         90        100
CCATCATTAC CGCGCGGGCA TGGCGATGGG CACCACGTTG CAGGTGCCGC

         110        120        130        140        150
CGCAGATGTG GCCACCGGAT CGGGCGGCCT TCGACCGCTA CTGGCGGCAA

         160        170        180        190        200
TCACTGGACA GGGTGCACAT CGATGACGTC GTTCGCGACT ACCTGTATCC

         210        220        230        240        250
GATCGTGGCG CTCCGAATTC GCGGGATCGC ACTGCCGGGT CCGCTGCGGC

         260        270        280        290        300
GGCTGTCGGA GGGTATCGCG CTGCTGATCA CCACCGGTTT CCTGCCGCAG

         310        320        330        340        350
CGGTTTCGCG ACGAGATGCG GTTGCCGTGG GACGCGACCA AGCAGCGGCG

         360        370        380        390        400
CTTTGACGCG CTCATGGCCG TGCTGCGCAC GGTGAATCGC CTGATGCCGC

         410        420        430        440        450
GGTTTGTCCG GGAGTTCCCG TTCAACCTGA TGCTCTGGGA CCTGGACCGG

         460        470        480        490        500
CGGATGAGGC GCGGGCGCCC GCTGGTGTAA TCGACGGCTT CGCGTGGACC

         510        520        530        540        550
GATGGCGGTA GACCGCTCGC TAGATTGGCG GGCGAATTTG GTGCACAGAG

         560        570        580        590        600
GCAAACCGGG CGAAATCCCT ATCCAGGCTC ACCACGGCGC AGTGATGCTC

         610        620        630        640        650
CACGGCGATG GCCCCGAGTA CCGCGTCAGG TATCAAGTCG CCCGATGCGT

         660        670        680        690        700
CGGCCTCGTC GCAGAGTTTT CGCAGCAGCA CCAGGTGTCT GGGGCCGGGG

         710        720        730        740        750
CTTGTCGGAA GGTCATGGGG CTGGGCGTTG ACGGCTTCGA CGAATGCGAA

         760        770        780        790        800
TGCATCCGCT CGTGGTGACG GAATCTCGAA GATGCGTCGA TTCGTTGTTA

         810        820        830        840        850
GCCGGAGGAA CGACGCCCAC ACTAGGTTCG GCACTGTGAA GGGGTCGTCG

         860        870        880        890        900
GCCGCAAGCA GTCGATCGAA CCAGGGGCGG ACGGTTCGGT GATTCGGATG

         910        920        930        940        950
GTCACCGCGG TGTGCAGCCA GCAGCACGTT GACGTCGATG AGGAACATCG
```

Fig. 9
2/5

960 970 980 990 1000
CCTATTTGTG CCTGTCCAGG CTCACTTCCG CGAGTTCAGT TCCAGACCCT

1010 1020 1030 1040 1050
CGTCGAGCAC TTCGGACAAC ACCGTATTCG AGGTTAGGTC GATACCTGGC

1060 1070 1080 1090 1100
CGCGGACCGG TGCCGGCGTC AAAAACGGGG ACGGTTGGCC GGGCGCCGCC

1110 1120 1130 1140 1150
GGTACGGGCG GCGGCGAGCT CCCGCCGAAG GGCGTCTTCG ATCACAGCGC

1160 1170 1180 1190 1200
CCAGCGATTA ACCACGCTCG CGGGCCCGGC GTTTGGCGGT AGCCAGTAGT

1210 1220 1230 1240 1250
TCATCCGAGA TTGACACGGT GGTGCGCATG ATGCTCAGGA TAGCGCATCT

1260 1270 1280 1290 1300
ACGGCATCAT CTGCGGTGAG CAACTGATGC CCTCAACGCC GCGTGTGGTC

1310 1320 1330 1340 1350
GCAGGTCTGC CTGCTATGGC AAGCCGTTGA GTCCGTTCTC GCCGAGCAGC

1360 1370 1380 1390 1400
AGCCCGCCGG TGCCGCCGGC ACCGGGCGTG GCCCCGGCTT TGCCGGCGTT

1410 1420 1430 1440 1450
GGCGCCGTTG CCGCCGTTGC CGATCAGCAC GGCGTTGCCG CCGACACCAC

1460 1470 1480 1490 1500
CGCTGCCGCC GGTACCGGCG CCAAACCCGC CGGCAACCCC CGTCACCGCC

1510 1520 1530 1540 1550
GTTGCCGAAC ACCCCGGCGT GGCCACCGTC ACCGCCGGTG CCGCCGGTAC

1560 1570 1580 1590 1600
CGGCGCCTAG AGCGTTGGCA CCGCTGCCGC CGGCGCCGCC GGCGCCGGCG

1610 1620 1630 1640 1650
GAGCCGAAGA GCAAGCCGCC GTTCCCGCCG GCGCCGCCGG CGCCGCCTTG

1660 1670 1680 1690 1700
CTGGATGCTG GTAAGTGCTG CCCCGCCGTG CCCGCCGGCG CCGCCGGCGC

1710 1720 1730 1740 1750
CGCCGAAGCC GAAGAGTAAG GCGCCGTTCC CGCCGGTTCC GCCGGCCCCG

1760 1770 1780 1790 1800
CCGGCAAGGG AGCTGGCGCC ACCGCTGCCG CCGGCGCCAC CGGAGGCGCC

1810 1820 1830 1840 1850
GAGGGAGAGT AGGCCGGCGT TGCCGCCGTG CCCGCCGCCG CCGGTGGTGA

1860 1870 1880 1890 1900
TCCCGGACCC TCCCGAGCCG GCGGCGCCGC CGGTGCCGCC GGCTCCGAAC

Fig. 9 3/5

```
1910        1920        1930        1940        1950
AGTCCGCCGT TCCCGCCGTT CCCACCGGCC CCGAAGTTCG TGCCGGCCCC
1960        1970        1980        1990        2000
GCCGGTGCCG CCAGTTCCGA ACAGTCCGCC GTTCCCGCCG TTCCCGCCGG
2010        2020        2030        2040        2050
CTGCGTTGAA CCCGCCGGCC CCTCCGGCTC CGCCGTTGGC GAACAGTCCG
2060        2070        2080        2090        2100
CCGTTGCCGC CGGCGCCGCC GACGCCGGCC GGGACACCGC CAGCGGCGCC
2110        2120        2130        2140        2150
GTGGCCGCCG GTGCCGGCCG CGCCGAAGAG CAAACCGGCG TCGCCGCCGC
2160        2170        2180        2190        2200
GCCCGCCGGC CCCGCCGATG CCAGCGACGC CTATGGAGTT CCCACCGTTG
2210        2220        2230        2240        2250
CCGCCGGTGC CGCCGGATCC GATCAGCAAG GAGACCCCAC CGGCGCCGCC
2260        2270        2280        2290        2300
GGCCCCGCCG ATCCCTCCAG CACCGGTGCC TATCCCGCCG GTCCCGCCAT
2310        2320        2330        2340        2350
TGCCACCGGT ACCGAACAAG ATCCCGCCGG CCCCGCCGGC CCCGCCCGTA
2360        2370        2380        2390        2400
GCCGTGGCGG CGGTGTTGGT CGCACCGTGC CCGCCGTTAC CGCCGTTGCC
2410        2420        2430        2440        2450
GAACAACCAC CCGCCGGCCC CGCCGGCAGC CCCGGTCCCC GGGGTCCCGT
2460        2470        2480        2490        2500
TGGCGCCGTT GCCGAACAGC CACCCGCCGG CCCCGCCGTC AGCCCCGGTT
2510        2520        2530        2540        2550
CCAGGAGTCC CGTTGGCGCC GTTGCCGATC AGCGGGCGGC CGGTGAGCGT
2560        2570        2580        2590        2600
CTGGAAGGGC TCGTTCACCA CATTGAGCAC ATTTTGCTGC AGGGTGTGCA
2610        2620        2630        2640        2650
GTGGCGAGGT GCTCGCGGGA GCATTGAATC CGTCTAGACC GAGCAGGAGC
2660        2670        2680        2690        2700
CCGCTGACGA CGACCACTCC GGCCTTGCCC GCGCCAATCC CACCGCTACC
2710        2720        2730        2740        2750
GCCGTTACCG CCATTGCCGA TCAACACGGC GGTGCCACCG ATCCCGCCGT
2760        2770        2780        2790        2800
TGCCGCCGGT CACCGCGCTG GCGCCACCGT TACCGCCGTT GGCGCCGTTA
2810        2820        2830        2840        2850
CCGATCAGCC CGGGGGTGCC GCCAGCCCCA CCGATCCCGC CGGGGAAGCC
```

Fig. 9
4/5

```
          2860        2870        2880        2890        2900
CTGGACAACT CCGCCGTTGG CGCCGGCGCC GCCGGAGCCG AAGACCGTGC

          2910        2920        2930        2940        2950
CGGTGTTGCC CCCGGGGCCG TCTTGCCCGC CGTCGGAGAA GCCGAATCCG

          2960        2970        2980        2990        3000
CCGGCGCCGC CGGAGCCGCC GGAGCCGAAG AGCAGCCCAG CGTTGCCGCC

          3010        3020        3030        3040        3050
GGCGCCGCCG GCGCCGtCTA TGCCGTCGGC CGTGAGAGTA CCGCCGTCCC

          3060        3070        3080        3090        3100
CACCGATTCC GCCGGCGCCG CCCGCGGCGC CGAGGGCGAG CATGCCGGCA

          3110        3120        3130        3140        3150
TTGCCGCCGG CCCCGCCGTC CCCGCCGGCG ACCAGGCTGT GTCCGCCGCT

          3160        3170        3180        3190        3200
GCCGCCTTCC CCGCCTGCGC CGAACAGCCC GCCGGCCCCG CCGGCCCCGC

          3210        3220        3230        3240        3250
CGACTCCGCC GAAGCTGCTG TCGGCGAACC CGCCATGCCC GCCGGTGCCG

          3260        3270        3280        3290        3300
CCGGCGCCGA ACAGACCGCC AGCGCCACCG GCCCCACCGG CCCCGCCGGA

          3310        3320        3330        3340        3350
GCTGCCGGCC CCACCGGATC CGCCGACCCC GCCGGTGGCG AACAGCCCGC

          3360        3370        3380        3390        3400
CGGCCCCGCC GGCGCCGCCC GCCCCGCCGA GTGCACTGCC GTTCGTGAAT

          3410        3420        3430        3440        3450
CCGCCGGCCC CGCCGACTCC GGCGGCGCCG AAGAGCAGGC CGGCGTTGCC

          3460        3470        3480        3490        3500
GGCAGCCCCG CCGGCGCCGC CGGCCCCGCC CGTGAGGGCT ACTACGCCGC

          3510        3520        3530        3540        3550
CGCCGGCGCC GCCGGCGCCG CCGGCGCCGA ACAGCATGGC GTTGCCGCCG

          3560        3570        3580        3590        3600
GCTCCGCCGG ACCCGCCGAT CCCACTGCTG GCGACCCCGC CAGCGCCGCC

          3610        3620        3630        3640        3650
GGCGCCGCCG TTGCCGATGA GCCCGCCGGC GCCGCCGTTG CCGCCGGCCG

          3660        3670        3680        3690        3700
CGCCGGATCC TCCGGCGCCG CCGTTGACGA TTAACCAGCC GCCGTCCCCG

          3710        3720        3730        3740        3750
CCATTGGCCC CGGTGCCGGG GGCGCCGTTG GCGCCGTTGC CGATCAACGG

          3760        3770        3780        3790        3800
GCGCCCGGTA TTCGCCAGGA AGAACTCGTT GATCGGATCC AGCAGCGGCG
```

Fig. 9

5/5

```
        3810        3820        3830        3840        3850
ACACCGCGGC GGCCTCGGCG GCCGCATAGG CGCCGCCACC GGAGGTCAAT

        3860        3870        3880        3890        3900
GCCTGCACGA ACTGGGCATG AAACGCCTGC GCTTGGGCGC TGAGCGCCTG

        3910        3920        3930        3940        3950
ATAGGCCTGG CCGTGGGCGC CGAACAGCGC GGCGATGGCT GTCGAC
```

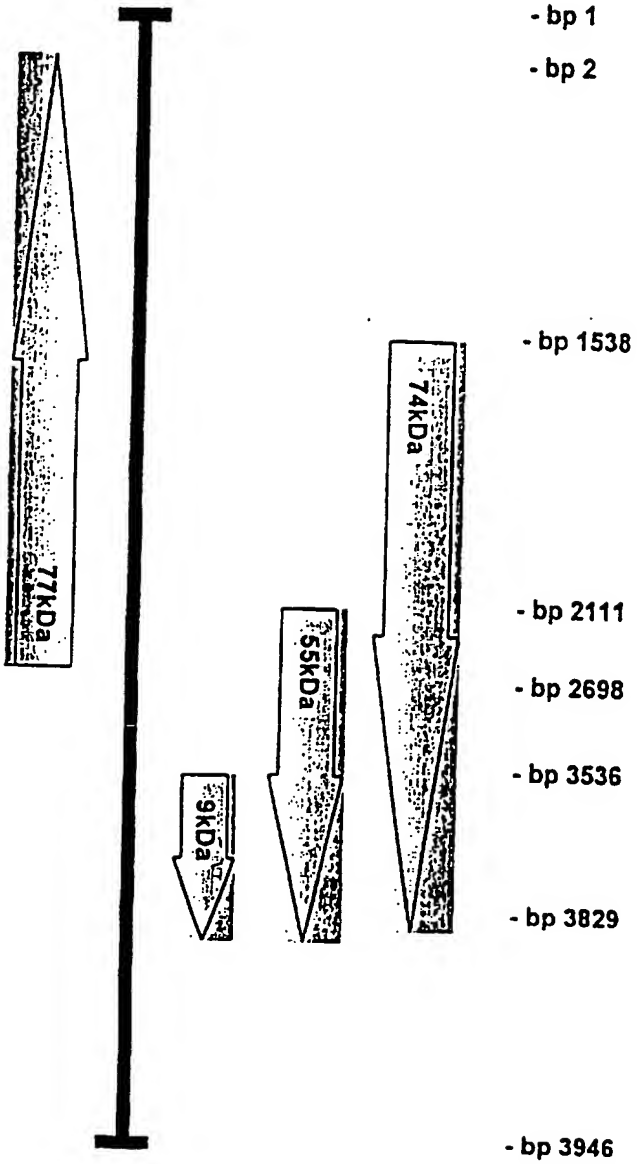


Figure 10. Schematic drawing of the clone Mtub-Clara-Klon. The open reading frames for about 9kDa (bp 3536 - bp 3829), 55kDa (bp 2111- bp 3829), 74 kDa (bp 1538 - bp 3829) and 77 kDa (bp 2698 - bp 2 on the complimentary strand) proteins are shown by arrows and the corresponding coding regions are numbered.

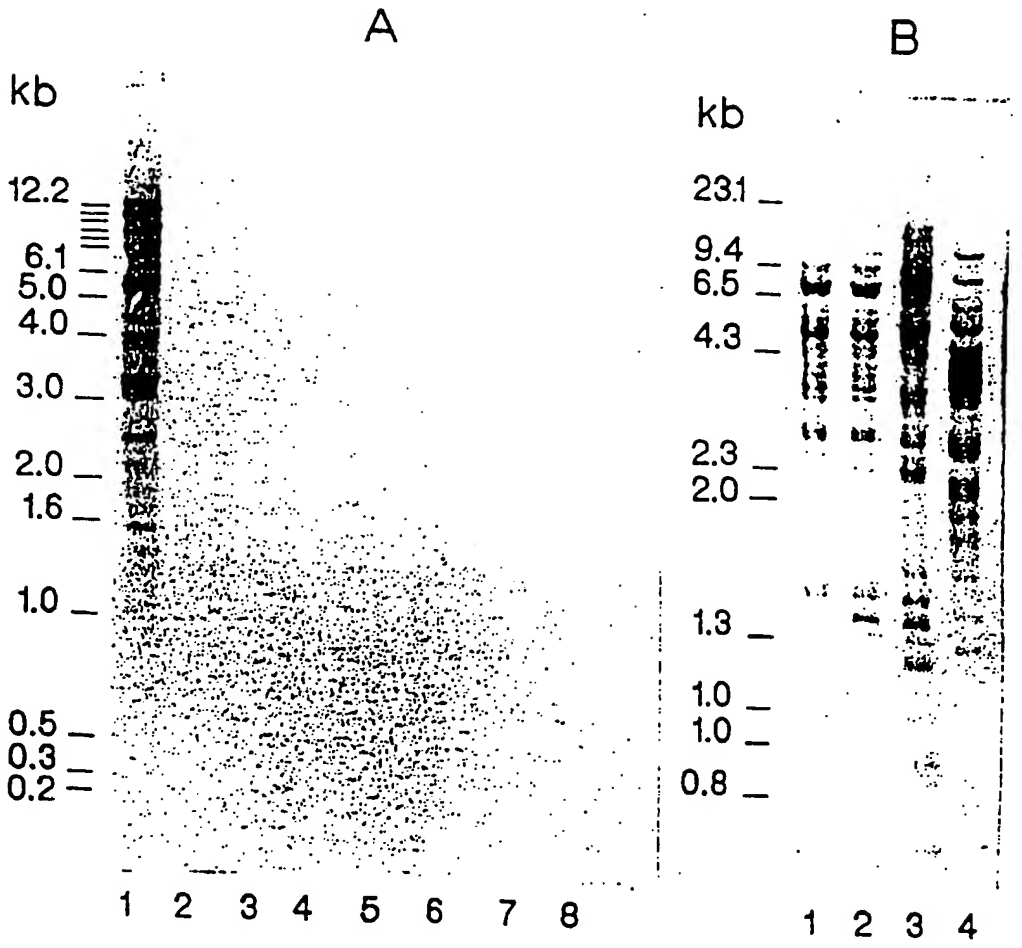


Fig. 11 A. Southern hybridization with genomic DNA from different mycobacteria digested with PvuII (1. M. tuberculosis H37Rv, 2. M. avium, 3. M. kanssasi, 4. M. necroti, 5. M.fortuitum, 6. M. phlei, 7. M. smegmatis, 8. M. vaccae.)

Fig. 11B
Finger-print obtained using the DNA (BamHI digest) of 1.M. tuberculosis H37 RV , 2. M. tuberculosis H37 Ra, 3. M. bovis BCG, and 4. M. tuberculosis H37Rv digested with Sal I.

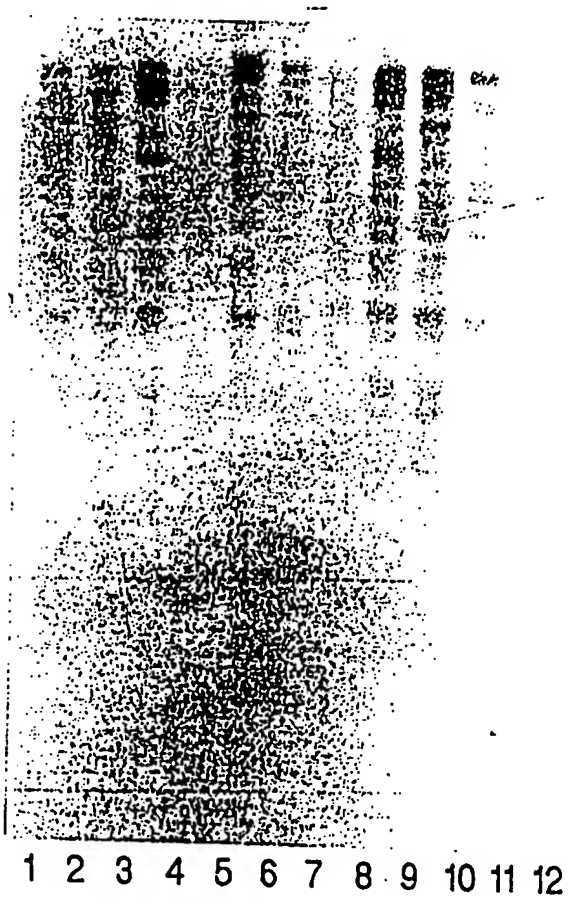


Fig 12. Finger-print with DNA fron differents M. tuberculosis clinical isolates (numbered 1- 12) digested with PvuII restricion enzyme. The 4 Kb Sal I fragment (Mtub-Klar-Klon) was used as probe.

Fig. 13 20/23

Amino acid sequence of the protein of about 74kDa.
Molecular-weight 74999 ; Length 764
THE AMINO ACID SEQUENCE IS GIVEN BELOW:

```
        10         20         30         40         50         60
VPPVPAPRAL APLPPAPPAP AEPKSKPPFP PAPPAPPCWM LVSAAPPCPP APPAPPKPKS
        70         80         90        100        110        120
KAPFPPVPPA PPARELAPPL PPAPPEAPRE SRPALPPCPP PPVVIPDPPE PAAPPVPPAP
       130        140        150        160        170        180
NSPPFPPFPP APKFVPAPPV PPVPNSPPFP PFPPAALNPP APPAPPLANS PPLPPAPPTP
       190        200        210        220        230        240
AGTPPAAPWP PVPAAPKSKP ASPPRPPAPP MPATPMEFPP LPPVPPDPIS KETPPAPPAP
       250        260        270        280        290        300
PIPPAPVPIP PVPPLPPVPN KIPPAPPAPP VAVAAVLVAP CPPLPPLPNN HPPAPPAAPV
       310        320        330        340        350        360
PGVPLAPLPN SHPPAPPSAP VPGVPLAPLP ISGRPVSVWK GSFTTLSTFC CRVCSGEVLA
       370        380        390        400        410        420
GALNPSRPSR SPLTTTTPAL PAPIPPLPPL PPLPINTAVP PIPPLPPVTA LAPPLPPLAP
       430        440        450        460        470        480
LPISPGVPPA PPIPPGKPWT TPPLAPAPPE PKTVPVLPPG PSCPPSEKPN PPAPPEPPEP
       490        500        510        520        530        540
KSSPALPPAP PAPSMPSAVR VPPSPPIPPA PPAAPRASMP ALPPAPPSPP ATRLCPPLPP
       550        560        570        580        590        600
SPPAPNSPPA PPAPPTPPKL LSANPPCPPV PPAPNRPPAP PAPPAPPELP APPDPPTPPV
       610        620        630        640        650        660
ANSPPAPPAP PAPPSALPFV NPPAPPTPAA PKSRPALPAA PPAPPAPPVR ATTPPPAPPA
       670        680        690        700        710        720
PPAPNSMALP PAPPDPPIPL LATPPAPPAP PLPMSPPAPP LPPAAPDPPA PPLTINQPPS
       730        740        750        760        770        780
PPLAPVPGAP LAPLPINGRP VFARKNSLIG SSSGDTAAAS AAA*...... ..........
```

Fig. 14

A glycine rich protein of about 77kDa.
Molecular-weight 77056 ; Length 899
THE AMINO ACID SEQUENCE IS GIVEN BELOW:

```
        10         20         30         40         50         60
STAIAALFGA HGQAYQALSA QAQAFHAQFV QALTSGGGAY AAAEAAAVSP LLDPINEFFL
        70         80         90        100        110        120
ANTGRPLIGN GANGAPGTGA NGGDGGWLIV NGGAGGSGAA GGNGGAGGLI GNGGAGGAGG
       130        140        150        160        170        180
VASSGIGGSG GAGGNAMLFG AGGAGGAGGG VVALTGGAGG AGGAAGNAGL LFGAAGVGGA
       190        200        210        220        230        240
GGFTNGSALG GAGGAGGAGG LFATGGVGGS GGAGSSGGAG GAGGAGGLFG AGGTGGHGGF
       250        260        270        280        290        300
ADSSFGGVGG AGGAGGLFGA GGEGGSGGHS LVAGGDGGAG GNAGMLALGA AGGAGGIGGD
       310        320        330        340        350        360
GGTLTADGID GAGGAGGNAG LLFGSGGSGG AGGFGFSDGG QDGPGGNTGT VFGSGGAGAN
       370        380        390        400        410        420
GGVVQGFPGG IGGAGGTPGL IGNGANGGNG GASAVTGGNG GIGGTAVLIG NGGNGGSGGI
       430        440        450        460        470        480
GAGKAGVVVV SGLLIGLDGF NAPASTSPLH TLQQNVLNVV NEPFQTLTGR PLIGNGANGT
       490        500        510        520        530        540
PGTGADGGAG GWLFGNGANG TPGTGAAGGA GGWLFGNGGN GGHGATNTAA TATGGAGGAG
       550        560        570        580        590        600
GILFGTGGNG GTGGIGTGAG GIGGAGGAGG VSLLIGSGGT GGNGGNSIGV AGIGGAGGRG
       610        620        630        640        650        660
GDAGLLFGAA GTGGHGAAGG VPAGVGGAGG NGGLFANGGA GGAGGFNAAG GNGGNGGLFG
       670        680        690        700        710        720
TGGTGGAGTN FGAGGNGGNG GLFGAGGTGG AAGSGGSGIT TGGGGHGGNA GLLSLGASGG
       730        740        750        760        770        780
AGGSGGASSL AGGAGGTGGN GALLFGFGGA GGAGGHGGAA LTSIQQGGAG GAGGNGGLLF
       790        800        810        820        830        840
GSAGAGGAGG SGANALGAGT GGTGGDGGHA GVFGNGGDGG CRRVWRRYRR QRWCRRQRRA
       850        860        870        880        890        900
DRQRRQRRQR RQSRGHARCR RHRRAAARRE RTQRLAIAGR PATTRGVEGI SCSPQMMP*.
```

Fig. 15 22/23

Amino acid sequence of the about 9 kDa proline rich protein
Molecular-weight 9356 ; Length 98
AMINO ACID SEQUENCE IS GIVEN BELOW:

```
        10         20         30         40         50         60
MALPPAPPDP PIPLLATPPA PPAPPLPMSP PAPPLPPAAP DPPAPPLTIN QPPSPPLAPV
        70         80         90        100        110        120
PGAPLAPLPI NGRPVFARKN SLIGSSSGDT AAASAAA*.. .......... ..........
```

Fig. 16

Proline rich protein of about 55kDa.
Molecular-weight 55982 ; Length 573
AMINO ACID SEQUENCE IS GIVEN BELOW.

```
        10         20         30         40         50         60
VPAAPKSKPA SPPRPPAPPM PATPMEFPPL PPVPPDPISK ETPPAPPAPP IPPAPVPIPP
        70         80         90        100        110        120
VPPLPPVPNK IPPAPPAPPV AVAAVLVAPC PPLPPLPNNH PPAPPAAPVP GVPLAPLPNS
       130        140        150        160        170        180
HPPAPPSAPV PGVPLAPLPI SGRPVSVWKG SFTTLSTFCC RVCSGEVLAG ALNPSRPSRS
       190        200        210        220        230        240
PLTTTTPALP APIPPLPPLP PLPINTAVPP IPPLPPVTAL APPLPPLAPL PISPGVPPAP
       250        260        270        280        290        300
PIPPGKPWTT PPLAPAPPEP KTVPVLPPGP SCPPSEKPNP PAPPEPPEPK SSPALPPAPP
       310        320        330        340        350        360
APSMPSAVRV PPSPPIPPAP PAAPRASMPA LPPAPPSPPA TRLCPPLPPS PPAPNSPPAP
       370        380        390        400        410        420
PAPPTPPKLL SANPPCPPVP PAPNRPPAPP APPAPPELPA PPDPPTPPVA NSPPAPPAPP
       430        440        450        460        470        480
APPSALPFVN PPAPPTPAAP KSRPALPAAP PAPPAPPVRA TTPPPAPPAP PAPNSMALPP
       490        500        510        520        530        540
APPDPPIPLL ATPPAPPAPP LPMSPPAPPL PPAAPDPPAP PLTINQPPSP PLAPVPGAPL
       550        560        570        580        590        600
APLPINGRPV FARKNSLIGS SSGDTAAASA AA*....... .......... ..........
```